APPLICATION

FOR

UNITED STATES LETTERS PATENT

TITLE: ANDROGEN RECEPTOR COMPLEX-ASSOCIATED
              PROTEIN

APPLICANT: TAI-JAY CHANG

ANDROGEN RECEPTOR COMPLEX-ASSOCIATED PROTEIN

Cross-Reference to Related Applications

This application claims priority from U.S. Provisional Application No. 60/262,312,

5   filed January 17, 2001.

Background of the Invention

A variety of genes that are overexpressed in tumor cells relative to healthy cells have

been identified. It is expected that the identification of such genes will provide drug targets

10   for anti-cancer drug development and for cancer diagnostics. The number of steroid

receptors (e.g., androgen receptors) in liver tumors cells appears to be increased relative to

their adjacent healthy liver cells.

Steroid hormones generally exert their physiological effects by binding to their

specific nuclear receptors to form complexes that in turn act as transcription factors. The

15   complexes bind to specific nucleotide sequences (steroid responsive elements) in the

promoters of steroid-responsive genes to facilitate transcription of those genes.

Summary of the Invention

The invention is based on the discovery of a human protein that is overexpressed in

20   hepatoma cells relative to normal adjacent tissue in liver cancer patients. It was also

discovered that this human protein binds to an androgen receptor and moreover augments the

ability of the androgen receptor to transactivate an androgen-responsive gene. Thus, the

human protein to which this invention pertains was designated androgen receptor complex-

associated protein or ARCAP. The full-length human ARCAP cDNA, with the start and stop

25   codons underlined, is shown below.

```
     CCGGCTCAGGCAGAGCCATGTCTCGGGGTGGCTCCTACCCACACCTGTTGTGGGACGTGAGGAAAAGGTCCCTC
     GGGCTGGAGGACCCGTCCCGGCTGCGGAGTCGCTACCTGGGAAGAAGAGAATTTATCCAAAGATTAAAACTTGA
     AGCAACCCTTAATGTGCATGATGGTTGTGTTAATACAATCTGTTGGAATGACACTGGAGAATATATTTTATCTG
30   GCTCAGATGACACCAAATTAGTAATTAGTAATCCTTACAGCAGAAAGGTTTTGACAACAATTCGTTCAGGGCAC
     CGAGCAAACATATTTAGTGCAAAGTTCTTACCTTGTACAAATGATAAACAGATTGTATCCTGCTCTGGAGATGG
     AGTAATATTTTATACCAACGTTGAGCAAGATGCAGAAACCAACAGACAATGCCAATTTACGTGTCATTATGGAA
     CTACTTATGAGATTATGACTGTACCCAATGACCCTTACACTTTTCTCTCTTGTGGTGAAGATGGAACTGTTAGG
     TGGTTTGATACACGCATCAAAACTAGCTGCACAAAAGAAGATTGTAAAGATGATATTTTAATTAACTGTCGACG
35   TGCTGCCACGTCTGTTGCTATTTGCCCACCAATACCATATTACCTTGCTGTTGGTTGTTCTGACAGCTCAGTAC
```

```
GAATATATGATCGGCGAATGCTGGGCACAAGAGCTACAGGGAATTATGCAGGTCGAGGGACTACTGGAATGGTT
GCCCGTTTTATTCCTTCCCATCTTAATAATAAGTCCTGCAGAGTGACATCTCTGTGTTACAGTGAAGATGGTCA
AGAGATTCTCGTTAGTTACTCTTCAGATTACATATATCTTTTTGACCCGAAAGATGATACAGCACGAGAACTTA
AAACTCCTTCTGCGGAAGAGAGAAGAGAAGAGTTGCGACAACCACCAGTTAAGCGTTTGAGACTTCGTGGTGAT
```
5
```
TGGTCAGATACTGGACCCAGAGCAAGGCCGGAGAGTGAACGAGAACGAGATGGAGAGCAGAGTCCCAATGTGTC
ATTGATGCAGAGAATGTCTGATATGTTATCAAGATGGTTTGAAGAAGCAAGTGAGGTTGCACAAAGCAATAGAG
GACGAGGAAGATCTCGACCCAGAGGTGGAACAAGTCAATCAGATATTTCAACTCTTCCTACGGTCCCATCAAGT
CCTGATTTGGAAGTGAGTGAAACTGCAATGGAAGTAGATACTCCAGCTGAACAATTTCTTCAGCCTTCTACATC
CTCTACAATGTCAGCTCAGGCTCATTCGACATCATCTCCCACAGAAAGCCCTCATTCTACTCCTTTGCTATCTT
```
10
```
CTCCAGACAGTGAACAAAGGCAGTCTGTTGAGGCATCTGGACACCACACACATCATCAGTCTGATAACAATAAT
GAAAAGCTGAGCCCCAAACCAGGGACAGGTGAACCAGTTTTAAGTTTGCACTACAGCACAGAAGGAACAACTAC
AAGCACAATAAAACTGAACTTTACAGATGAATGGAGCAGTATAGCATCAAGTTCTAGAGGAATTGGGAGCCATT
GCAAATCTGAGGGTCAGGAGGAATCTTTCGTCCCACAGAGCTCAGTGCAACCACCAGAAGGAGACAGTGAAACA
AAAGCTCCTGAAGAATCATCAGAGGATGTGACAAAATATCAGGAAGGAGTATCTGCAGAAAACCCAGTTGAGAA
```
15
```
CCATATCAATATAACACAATCAGATAAGTTCACAGCCAAGCCATTGGATTCCAACTCAGGAGAAAGAAATGACC
TCAATCTTGATCGCTCTTGTGGGGTTCCAGAAGAATCTGCTTCATCTGAAAAAGCCAAGGAACCAGAAACTTCA
GATCAGACTAGCACTGAGAGTGCTACCAATGAAAATAACACCAATCCTGAGCCTCAGTTCCAAACAGAAGCCAC
TGGGCCTTCAGCTCATGAAGAAACATCCACCAGGGACTCTGCTCTTCAGGACACAGATGACAGTGATGATGACC
CAGTCCTGATCCCAGGTGCAAGGTATCGAGCAGGACCTGGTGATAGACGCTCTGCTGTTGCCCGTATTCAGGAG
```
20
```
TTCTTCAGACGGAGAAAAGAAAGGAAAGAAATGGAAGAATTGGATACTTTGAACATTAGAAGGCCGCTAGTAAA
AATGGTTTATAAAGGCCATCGCAACTCCAGGACAATGATAAAAGAAGCCAATTTCTGGGGTGCTAACTTTGTAA
TGAGTGGTTCTGACTGTGGCCACATTTTCATCTGGGATCGGCACACTGCTGAGCATTTGATGCTTCTGGAAGCT
GATAATCATGTGGTAAACTGCCTGCAGCCACATCCGTTTGACCCAATTTTAGCCTCATCTGGCATAGATTATGA
CATAAAGATCTGGTCACCATTAGAAGAGTCAAGGATTTTTAACCGAAAACTTGCTGATGAAGTTATAACTCGAA
```
25
```
ACGAACTCATGCTGGAAGAAACTAGAAACACCATTACAGTTCCAGCCTCTTTCATGTTGAGGATGTTGGCTTCA
CTTAATCATATCCGAGCTGACCGGTTGGAGGGTGACAGATCAGAAGGCTCTGGTCAAGAGAATGAAAATGAGGA
TGAGGAATAATAAACTCTTTTTGGCAAGCACTTAAATGTTCTGAAATTTGTATAAGACATTTATTATATTTTTT
TCTTTACAGAGCTTTAGTGCAATTTTAAGGTTATGGTTTTTGGAGTTTTTCCCTTTTTTTGGGATAACCTAACA
TTGGTTTGGAATGATTGTGTGCATGAATTTGGGAGATTGTATAAAACAAAACTAGCAGAATGTTTTTAAAACTT
```
30
```
TTTGCCGTGTATGAGGAGTGCTAGAAAATGCAAAGTGCAATATTTTCCCTAACCTTCAAATGTGGGAGCTTGGA
TCAATGTTGAAGAATAATTTTCATCATAGTGAAAATGTTGGTTCAAATAAATTTCTACACTTGCCATTTGCATG
TTTGTTGCTTTCTAATTAAAGAAACTGGTTGTTTAAAAAAAAAAAAAAGGAATTC (SEQ ID NO:3)
```

The nucleotide sequence encoding the human ARCAP protein (i.e., from the ATG start codon

35 to the codon immediately before the stop codon in SEQ ID NO:3) is designated SEQ ID

NO:1. The ARCAP amino acid sequence encoded by the above cDNA is shown below.

```
Met ser arg gly gly ser tyr pro his leu leu trp asp val arg lys arg ser
leu gly leu glu asp pro ser arg leu arg ser arg tyr leu gly arg arg glu
phe ile gln arg leu lys leu glu ala thr leu asn val his asp gly cys val
asn thr ile cys trp asn asp thr gly glu tyr ile leu ser gly ser asp asp
thr lys leu val ile ser asn pro tyr ser arg lys val leu thr thr ile arg
ser gly his arg ala asn ile phe ser ala lys phe leu pro cys thr asn asp
lys gln ile val ser cys ser gly asp gly val ile phe tyr thr asn val glu
gln asp ala glu thr asn arg gln cys gln phe thr cys his tyr gly thr thr
tyr glu ile met thr val pro asn asp pro tyr thr phe leu ser cys gly glu
asp gly thr val arg trp phe asp thr arg ile lys thr ser cys thr lys glu
asp cys lys asp asp ile leu ile asn cys arg arg ala ala thr ser val ala
ile cys pro pro ile pro tyr tyr leu ala val gly cys ser asp ser ser val
arg ile tyr asp arg arg met leu gly thr arg ala thr gly asn tyr ala gly
arg gly thr thr gly met val ala arg phe ile pro ser his leu asn asn lys
ser cys arg val thr ser leu cys tyr ser glu asp gly gln glu ile leu val
ser tyr ser ser asp tyr ile tyr leu phe asp pro lys asp asp thr ala arg
```

```
     glu leu lys thr pro ser ala glu glu arg arg glu glu leu arg gln pro pro
     val lys arg leu arg leu arg gly asp trp ser asp thr gly pro arg ala arg
     pro glu ser glu arg glu arg asp gly glu gln ser pro asn val ser leu met
     gln arg met ser asp met leu ser arg trp phe glu glu ala ser glu val ala
  5  gln ser asn arg gly arg gly arg ser arg pro arg gly gly thr ser gln ser
     asp ile ser thr leu pro thr val pro ser ser pro asp leu glu val ser glu
     thr ala met glu val asp thr pro ala glu gln phe leu gln pro ser thr ser
     ser thr met ser ala gln ala his ser thr ser ser pro thr glu ser pro his
     ser thr pro leu leu ser ser pro asp ser glu gln arg gln ser val glu ala
 10  ser gly his his thr his his gln ser asp asn asn asn glu lys leu ser pro
     lys pro gly thr gly glu pro val leu ser leu his tyr ser thr glu gly thr
     thr thr ser thr ile lys leu asn phe thr asp glu trp ser ser ile ala ser
     ser ser arg gly ile gly ser his cys lys ser glu gly gln glu glu ser phe
     val pro gln ser ser val gln pro pro glu gly asp ser glu thr lys ala pro
 15  glu glu ser ser glu asp val thr lys tyr gln glu gly val ser ala glu asn
     pro val glu asn his ile asn ile thr gln ser asp lys phe thr ala lys pro
     leu asp ser asn ser gly glu arg asn asp leu asn leu asp arg ser cys gly
     val pro glu glu ser ala ser ser glu lys ala lys glu pro glu thr ser asp
     gln thr ser thr glu ser ala thr asn glu asn asn thr asn pro glu pro gln
 20  phe gln thr glu ala thr gly pro ser ala his glu glu thr ser thr arg asp
     ser ala leu gln asp thr asp asp ser asp asp asp pro val leu ile pro gly
     ala arg tyr arg ala gly pro gly asp arg arg ser ala val ala arg ile gln
     glu phe phe arg arg arg lys glu arg lys glu met glu glu leu asp thr leu
     asn ile arg arg pro leu val lys met val tyr lys gly his arg asn ser arg
 25  thr met ile lys glu ala asn phe trp gly ala asn phe val met ser gly ser
     asp cys gly his ile phe ile trp asp arg his thr ala glu his leu met leu
     leu glu ala asp asn his val val asn cys leu gln pro his pro phe asp pro
     ile leu ala ser ser gly ile asp tyr asp ile lys ile trp ser pro leu glu
     glu ser arg ile phe asn arg lys leu ala asp glu val ile thr arg asn glu
 30  leu met leu glu glu thr arg asn thr ile thr val pro ala ser phe met leu
     arg met leu ala ser leu asn his ile arg ala asp arg leu glu gly asp arg
     ser glu gly ser gly gln glu asn glu asn glu asp glu glu (SEQ ID NO:2)
```

Accordingly, the invention features a substantially pure polypeptide or protein

35 including an amino acid sequence at least 70% (e.g., at least 75, 80, 85, 90, 95, 98, or 100%)

identical to SEQ ID NO:2. If the polypeptide includes a sequence that is 100% identical to

SEQ ID NO:2, the polypeptide can contain up to 30 conservative amino acid substitutions.

The invention also includes a substantially pure polypeptide encoded by a nucleic acid that

hybridizes under stringent conditions to a probe the sequence of which consists of SEQ ID

40 NO:1. The polypeptide binds to an androgen receptor and increases the ability of the

androgen receptor to transactivate an androgen-responsive gene, as shown in the Example

below.

The invention further features an isolated nucleic acid encoding a polypeptide of the

invention, a vector including a nucleic acid of the invention, and a cultured host cell

45 containing a nucleic acid of the invention. An example of a nucleic acid within the invention

includes an isolated nucleic acid having a strand that hybridizes under stringent conditions to

3

a single stranded probe, the sequence of which consists of SEQ ID NO:1 or the complement of SEQ ID NO:1. Such a nucleic acid can be at least 15 (e.g., at least 30, 50, 100, 200, 500, or 1000) nucleotides in length.

5   In addition, the invention features a method of producing a polypeptide by culturing a cultured host cell of the invention in a culture, expressing the polypeptide in the cultured host cell, and isolating the polypeptide from the culture.

The invention also features a method of screening for a compound that decreases androgen receptor-mediated transactivation by contacting a polypeptide of the invention with a protein complex including an androgen receptor, in the presence of a candidate compound;

10   measuring the extent of binding between the polypeptide and the protein complex; and determining whether the extent of binding is less than the extent of binding between the polypeptide and the protein complex in the absence of the candidate compound. An extent of binding in the presence of the compound less than the extent of binding in the absence of the compound indicates that the candidate compound decreases androgen receptor-mediated

15   transactivation.

The term "substantially pure" as used herein in reference to a given polypeptide means that the polypeptide is substantially free from other biological macromolecules. The substantially pure polypeptide is at least 75% (e.g., at least 80, 85, 95, or 99%) pure by dry weight. Purity can be measured by any appropriate standard method, for example, by

20   column chromatography, polyacrylamide gel electrophoresis, or HPLC analysis.

A "conservative amino acid substitution" is one in which an amino acid residue is replaced with another residue having a chemically similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g.,

25   aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine).

30   By hybridization under "stringent conditions" is meant hybridization at 65°C, 0.5 X SSC, followed by washing at 45°C, 0.1 X SSC.

The "percent identity" of two amino acid sequences or of two nucleic acids is determined using the algorithm of Karlin and Altschul (Proc. Natl. Acad. Sci. USA 87:2264-2268, 1990), modified as in Karlin and Altschul (Proc. Natl. Acad. Sci. USA 90:5873-5877, 1993). Such an algorithm is incorporated into the NBLAST and XBLAST programs of

5   Altschul et al. (J. Mol. Biol. 215:403-410, 1990). BLAST nucleotide searches are performed with the NBLAST program, score = 100, wordlength = 12. BLAST protein searches are performed with the XBLAST program, score = 50, wordlength = 3. Where gaps exist between two sequences, Gapped BLAST is utilized as described in Altschul et al. (Nucleic Acids Res. 25:3389-3402, 1997). When utilizing BLAST and Gapped BLAST programs, the

10   default parameters of the respective programs (e.g., XBLAST and NBLAST) are used. See http://www.ncbi.nlm.nih.gov.

An "isolated nucleic acid" is a nucleic acid the structure of which is not identical to that of any naturally occurring nucleic acid or to that of any fragment of a naturally occurring genomic nucleic acid spanning more than three separate genes. The term therefore covers,

15   for example, (a) a DNA which has the sequence of part of a naturally occurring genomic DNA molecule but is not flanked by both of the coding sequences that flank that part of the molecule in the genome of the organism in which it naturally occurs; (b) a nucleic acid incorporated into a vector or into the genomic DNA of a prokaryote or eukaryote in a manner such that the resulting molecule is not identical to any naturally occurring vector or genomic

20   DNA; (c) a separate molecule such as a cDNA, a genomic fragment, a fragment produced by polymerase chain reaction (PCR), or a restriction fragment; and (d) a recombinant nucleotide sequence that is part of a hybrid gene, i.e., a gene encoding a fusion protein. Specifically excluded from this definition are nucleic acids present in mixtures of different (i) DNA molecules, (ii) transfected cells, or (iii) cell clones, e.g., as these occur in a DNA library such

25   as a cDNA or genomic DNA library.

The polypeptides of the invention can be used to generate antibodies (either monoclonal or polyclonal) that specifically bind to ARCAP protein. These antibodies in turn are useful for detecting the presence and distribution of ARCAP in tissues and in cellular compartments. For example, such antibodies can be used to diagnose cancerous liver tissue

30   by determining whether ARCAP protein is expressed or overexpressed in the tissue. Similarly, the nucleic acids of the invention can be used to diagnose liver cancer by

determining whether ARCAP mRNA is being expressed or overexpressed in a tissue or cell. The nucleic acids can be used as primers in PCR-based detection methods, or as labeled probes in nucleic acid blots (e.g., Northern blots).

5   Other features or advantages of the present invention will be apparent from the following detailed description, and also from the claims.

Detailed Description

        The invention relates to new ARCAP proteins and nucleic acids encoding them that are overexpressed in hepatocellular carcinoma cells relative to normal liver cells. In addition

10   to differential expression, ARCAP was found to bind to and augment the transactivation activity of an androgen receptor. These observations and others described below suggest that ARCAP activates, via an androgen receptor complex, mitogenic genes that are androgen-responsive (i.e., genes whose promoters contain androgen responsive elements), that overexpression of ARCAP leads to cancer by facilitating androgen receptor-mediated

15   transactivation of androgen-responsive mitogenic genes, and that inhibition of ARCAP expression or activity would reduce expression of these androgen-responsive mitogenic genes and revert cancer cells to a more normal phenotype. Consequently, ARCAP is a new cancer drug target.

Uses

20   The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (e.g., therapeutic and prophylactic).

        The isolated nucleic acid molecules of the invention can be used, for example, to

25   express an ARCAP protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect an ARCAP mRNA (e.g., in a biological sample) or a genetic alteration in an ARCAP gene, and to modulate ARCAP activity. The ARCAP proteins can be used to treat disorders characterized by insufficient or excessive production of an ARCAP substrate or production of ARCAP inhibitors. In addition, the ARCAP proteins can be used

30   to screen for naturally occurring ARCAP substrates, to screen for drugs or compounds which modulate ARCAP activity, as well as to treat disorders characterized by insufficient or

excessive production of ARCAP protein or production of ARCAP protein forms that have

decreased, aberrant, or unwanted activity compared to ARCAP wild type protein (e.g., in

liver cancer). Moreover, the anti-ARCAP antibodies of the invention can be used to detect

and isolate ARCAP proteins, regulate the bioavailability of ARCAP proteins, and modulate

5   ARCAP activity.

A method of evaluating a compound for the ability to interact with, e.g., bind, a

subject ARCAP polypeptide is provided. The method includes: contacting the compound

with the subject ARCAP polypeptide; and evaluating ability of the compound to interact

with, e.g., to bind or form a complex with, the subject ARCAP polypeptide. This method can

10   be performed *in vitro*, e.g., in a cell free system, or *in vivo*, e.g., in a two-hybrid interaction

trap assay. This method can be used to identify naturally occurring molecules that interact

with subject ARCAP polypeptide. It can also be used to find natural or synthetic inhibitors

of a subject ARCAP polypeptide.

Screening Assays

15   The invention provides methods (also referred to herein as "screening assays") for

identifying modulators, i.e., candidate or test compounds or agents (e.g., proteins, peptides,

peptidomimetics, peptoids, small molecules, or other drugs) which bind to ARCAP proteins,

have a stimulatory or inhibitory effect on, for example, ARCAP expression or ARCAP

activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity

20   of an ARCAP substrate. Compounds thus identified can be used to modulate the activity of

target gene products (e.g., ARCAP genes) in a therapeutic protocol, to elaborate the

biological function of the target gene product, or to identify compounds that disrupt normal

target gene interactions.

In one embodiment, the invention provides assays for screening candidate or test

25   compounds which are substrates of an ARCAP protein or polypeptide or a biologically active

portion thereof. In another embodiment, the invention provides assays for screening

candidate or test compounds which bind to or modulate the activity of an ARCAP protein or

polypeptide or a biologically active portion thereof.

The test compounds of the present invention can be obtained using any of the

30   numerous approaches in combinatorial library methods known in the art, including:

biological libraries; peptoid libraries (libraries of molecules having the functionalities of

peptides, but with a novel, non-peptide backbone which is resistant to enzymatic degradation but which nevertheless remains bioactive; see, e.g., Zuckermann, R.N. *et al.* (1994) *J. Med. Chem.* 37:2678-85); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library

5   method; and synthetic library methods using affinity chromatography selection. The biological library and peptoid library approaches are limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer, or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art,

10   for example in DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994). *J. Med. Chem.* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.* 37:1233.

15   Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Biotechniques* 13:412-421), on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner, USP 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc Natl Acad Sci USA* 89:1865-1869), or on phage (Scott and Smith (1990) *Science* 249:386-390; Devlin (1990) *Science* 249:404-406; Cwirla *et al.* (1990) *Proc.*

20   *Natl. Acad. Sci.* 87:6378-6382; Felici (1991) *J. Mol. Biol.* 222:301-310; Ladner *supra.*).

In one embodiment, an assay is a cell-based assay in which a cell which expresses an ARCAP protein or biologically active portion thereof is contacted with a test compound, and the ability of the test compound to modulate ARCAP activity is determined. Determining the ability of the test compound to modulate ARCAP activity can be accomplished by

25   monitoring, for example, cell cycle-regulated cellular localization. The cell, for example, can be of mammalian origin, e.g., human.

The ability of the test compound to modulate ARCAP binding to a compound, e.g., an androgen receptor complex, or to bind to ARCAP can also be evaluated. This can be accomplished, for example, by coupling the compound, e.g., the substrate, with a

30   radioisotope or enzymatic label such that binding of the compound, e.g., the substrate, to ARCAP can be determined by detecting the labeled compound, e.g., substrate, in a complex.

Alternatively, ARCAP could be coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to modulate ARCAP binding to an ARCAP substrate in a complex. For example, compounds (e.g., ARCAP substrates) can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of

5   radioemmission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

The ability of a compound (e.g., an ARCAP substrate) to interact with ARCAP with

10   or without the labeling of any of the interactants can be evaluated. For example, a microphysiometer can be used to detect the interaction of a compound with ARCAP without the labeling of either the compound or the ARCAP. McConnell, H. M. *et al.* (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (e.g., Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-

15   addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and ARCAP.

In yet another embodiment, a cell-free assay is provided in which an ARCAP protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the ARCAP protein or biologically active portion thereof is

20   evaluated. Preferred biologically active portions of the ARCAP proteins to be used in assays of the present invention include fragments which participate in interactions with non-ARCAP molecules, e.g., fragments with high surface probability scores.

Soluble and/or membrane-bound forms of isolated proteins (e.g., ARCAP proteins or biologically active portions thereof) can be used in the cell-free assays of the invention.

25   When membrane-bound forms of the protein are used, it may be desirable to utilize a solubilizing agent. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane

30   sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate.

Cell-free assays involve preparing a reaction mixture of the target gene protein and the test compound under conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex that can be removed and/or detected.

The interaction between two molecules can also be detected, e.g., using fluorescence energy transfer (FET) (see, for example, Lakowicz et al., U.S. Patent No. 5,631,169; Stavrianopoulos, et al., U.S. Patent No. 4,868,103). A fluorophore label on the first, 'donor' molecule is selected such that its emitted fluorescent energy will be absorbed by a fluorescent label on a second, 'acceptor' molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the 'donor' protein molecule may simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the 'acceptor' molecule label may be differentiated from that of the 'donor.' Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the 'acceptor' molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (e.g., using a fluorimeter).

In another embodiment, determining the ability of the ARCAP protein to bind to a target molecule can be accomplished using real-time Biomolecular Interaction Analysis (BIA) (see, e.g., Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705). "Surface plasmon resonance" or "BIA" detects biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal which can be used as an indication of real-time reactions between biological molecules.

In one embodiment, the target gene product or the test substance is anchored onto a solid phase. The target gene product/test compound complexes anchored on the solid phase can be detected at the end of the reaction. Preferably, the target gene product can be anchored onto a solid surface, and the test compound, (which is not anchored), can be labeled, either directly or indirectly, with detectable labels discussed herein.

It may be desirable to immobilize either ARCAP, an anti-ARCAP antibody or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to an ARCAP protein, or interaction of an ARCAP protein with a target molecule

5   in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/ARCAP fusion proteins or glutathione-S-

10   transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or ARCAP protein, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation,

15   the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads. Complexes are determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of ARCAP binding or activity determined using standard techniques.

20   Other techniques for immobilizing either an ARCAP protein or a target molecule on matrices include using conjugation of biotin and streptavidin. Biotinylated ARCAP protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical).

25   In order to conduct the assay, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (e.g., by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously non-

30   immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is

not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the immobilized component (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody).

5   In one embodiment, this assay is performed utilizing antibodies reactive with ARCAP protein or target molecules but which do not interfere with binding of the ARCAP protein to its target molecule. Such antibodies can be derivatized to the wells of the plate, and unbound target or ARCAP protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the

10   ARCAP protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the ARCAP protein or target molecule.

Alternatively, cell free assays can be conducted in a liquid phase. In such an assay, the reaction products are separated from unreacted components by any of a number of standard techniques including but not limited to differential centrifugation (see, for example,

15   Rivas, G., and Minton, A.P., (1993) *Trends Biochem Sci* 18:284-7); chromatography (gel filtration chromatography, ion-exchange chromatography); electrophoresis (see, e.g., Ausubel, F. *et al.*, eds. Current Protocols in Molecular Biology 1999, J. Wiley: New York.); and immunoprecipitation (see, for example, Ausubel, F. *et al.*, eds. (1999) *Current Protocols in Molecular Biology*, J. Wiley: New York). Such resins and chromatographic techniques are

20   known to one skilled in the art (see, e.g., Heegaard, N.H., (1998) *J Mol Recognit* 11:141-8; Hage, D.S., and Tweed, S.A. (1997) *J Chromatogr B Biomed Sci Appl.* 699:499-525). Further, fluorescence energy transfer may also be conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

In a preferred embodiment, the assay includes contacting the ARCAP protein or

25   biologically active portion thereof with a known compound (e.g., an androgen receptor) which binds ARCAP to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an ARCAP protein, where determining the ability of the test compound to interact with an ARCAP protein includes determining the ability of the test compound to preferentially bind to

30   ARCAP or biologically active portion thereof, or to modulate the activity of a target molecule, as compared to the known compound.

The target gene products of the invention can, *in vivo*, interact with one or more cellular or extracellular macromolecules, such as proteins. For the purposes of this discussion, such cellular and extracellular macromolecules are referred to herein as "binding partners." Compounds that disrupt such interactions can be useful in regulating the activity

5   of the target gene product. Such compounds can include, but are not limited to molecules such as antibodies, peptides, and small molecules. The preferred target genes/products for use in this embodiment are the ARCAP genes herein identified. In an alternative embodiment, the invention provides methods for determining the ability of the test compound to modulate the activity of an ARCAP protein through modulation of the activity

10 of a downstream effector of an ARCAP target molecule. For example, the activity of the effector molecule on an appropriate target can be determined, or the binding of the effector to an appropriate target can be determined, as previously described.

To identify compounds that interfere with the interaction between the target gene product and its cellular or extracellular binding partner(s), a reaction mixture containing the

15 target gene product and the binding partner is prepared, under conditions and for a time sufficient, to allow the two products to form a complex. In order to test an inhibitory agent, the reaction mixture is provided in the presence and absence of the test compound. The test compound can be initially included in the reaction mixture, or can be added at a time subsequent to the addition of the target gene and its cellular or extracellular binding partner.

20 Control reaction mixtures are incubated without the test compound or with a placebo. The formation of any complexes between the target gene product and the cellular or extracellular binding partner is then detected. The formation of a complex in the control reaction, but not in the reaction mixture containing the test compound, indicates that the compound interferes with the interaction of the target gene product and the interactive binding partner.

25 Additionally, complex formation within reaction mixtures containing the test compound and normal target gene product can also be compared to complex formation within reaction mixtures containing the test compound and mutant target gene product. This comparison can be important in those cases where it is desirable to identify compounds that disrupt interactions of mutant but not normal target gene products.

30 These assays can be conducted in a heterogeneous or homogeneous format. Heterogeneous assays involve anchoring either the target gene product or the binding partner

onto a solid phase, and detecting complexes anchored on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the compounds being tested. For example, test compounds that interfere

5   with the interaction between the target gene products and the binding partners, e.g., by competition, can be identified by conducting the reaction in the presence of the test substance. Alternatively, test compounds that disrupt preformed complexes, e.g., compounds with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been

10   formed. The various formats are briefly described below.

In a heterogeneous assay system, either the target gene product or the interactive cellular or extracellular binding partner is anchored onto a solid surface (e.g., a microtiter plate), while the non-anchored species is labeled either directly or indirectly. The anchored species can be immobilized by non-covalent or covalent attachments. Alternatively, an

15   immobilized antibody specific for the species to be anchored can be used to anchor the species to the solid surface.

In order to conduct the assay, the partner of the immobilized species is exposed to the coated surface with or without the test compound. After the reaction is complete, unreacted components are removed (e.g., by washing) and any complexes formed will remain

20   immobilized on the solid surface. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the non-immobilized species is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface, e.g., using a labeled antibody specific for the initially non-immobilized species. The antibody, in turn, can be directly labeled or indirectly labeled

25   with, e.g., a labeled anti-Ig antibody. Depending upon the order of addition of reaction components, test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

Alternatively, the reaction can be conducted in a liquid phase in the presence or absence of the test compound, the reaction products separated from unreacted components,

30   and complexes detected, e.g., using an immobilized antibody specific for one of the binding components to anchor any complexes formed in solution and a labeled antibody specific for

the other partner to detect anchored complexes. Again, depending upon the order of addition of reactants to the liquid phase, test compounds that inhibit complex formation or that disrupt preformed complexes can be identified.

5   In an alternate embodiment, a homogeneous assay can be used. For example, a preformed complex of the target gene product and the interactive cellular or extracellular binding partner product is prepared so that either the target gene products or their binding partners are labeled, but the signal generated by the label is quenched due to complex formation (see, e.g., U.S. Patent No. 4,109,496 that utilizes this approach for immunoassays). The addition of a test substance that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way,

10   test substances that disrupt target gene product-binding partner interaction can be identified.

         In yet another aspect, the ARCAP proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993)

15   *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300) to identify other proteins, which bind to or interact with ARCAP ("ARCAP-binding proteins" or "ARCAP-bp") and are involved in ARCAP activity. Such ARCAP-bps can be activators or inhibitors of signals by the ARCAP proteins or ARCAP targets as, for example, downstream elements of an ARCAP-mediated signaling pathway.

20   The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for an ARCAP protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that

25   encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. (Alternatively, the ARCAP protein can be fused to the activator domain.) If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming an ARCAP-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription

30   of a reporter gene (e.g., lacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and

cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein that interacts with the ARCAP protein.

In another embodiment, modulators of ARCAP expression are identified. For example, a cell or cell free mixture is contacted with a candidate compound and the

5   expression of ARCAP mRNA or protein evaluated relative to the level of expression of ARCAP mRNA or protein in the absence of the candidate compound. When expression of ARCAP mRNA or protein is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of ARCAP mRNA or protein expression. Alternatively, when expression of ARCAP mRNA or protein is less (statistically

10   significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of ARCAP mRNA or protein expression. The level of ARCAP mRNA or protein expression can be determined by methods described herein for detecting ARCAP mRNA or protein.

In another aspect, the invention pertains to a combination of two or more of the

15   assays described herein. For example, a modulating agent can be identified using a cell-based or a cell free assay, and the ability of the agent to modulate the activity of an ARCAP protein can be confirmed *in vivo*, e.g., in an animal such as an animal model for hepatocellular carcinoma.

This invention further pertains to novel agents identified by the above-described

20   screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein (e.g., an ARCAP modulating agent, an antisense ARCAP nucleic acid molecule, an ARCAP-specific antibody, or an ARCAP-binding partner) in an appropriate animal model to determine the efficacy, toxicity, side effects, or mechanism of action, of treatment with such an agent. Furthermore, novel agents identified by the above-

25   described screening assays can be used for treating cancers, e.g., liver cancer.

Use of ARCAP Molecules as Surrogate Markers

The ARCAP molecules of the invention are also useful as markers of disorders or disease states, as markers for precursors of disease states, as markers for predisposition of disease states, as markers of drug activity, or as markers of the pharmacogenomic profile of a

30   subject. Using the methods described herein, the presence, absence and/or quantity of the ARCAP molecules of the invention may be detected, and may be correlated with one or more

biological states *in vivo*. For example, the ARCAP molecules of the invention may serve as surrogate markers for one or more disorders or disease states or for conditions leading up to disease states. As used herein, a "surrogate marker" is an objective biochemical marker which correlates with the absence or presence of a disease or disorder, or with the

5   progression of a disease or disorder (e.g., with the presence or absence of a liver tumor). The presence or quantity of such markers is independent of the disease. Therefore, these markers may serve to indicate whether a particular course of treatment is effective in lessening a disease state or disorder. Surrogate markers are of particular use when the presence or extent of a disease state or disorder is difficult to assess through standard methodologies (e.g., early

10   stage tumors), or when an assessment of disease progression is desired before a potentially dangerous clinical endpoint is reached (e.g., an assessment of cardiovascular disease may be made using cholesterol levels as a surrogate marker, and an analysis of HIV infection may be made using HIV RNA levels as a surrogate marker, well in advance of the undesirable clinical outcomes of myocardial infarction or fully-developed AIDS). Examples of the use of

15   surrogate markers in the art include those described in Koomen *et al.* (2000) *J. Mass. Spectrom.* 35: 258-264; and James (1994) *AIDS Treatment News Archive* 209.

The ARCAP molecules of the invention are also useful as pharmacodynamic markers. As used herein, a "pharmacodynamic marker" is an objective biochemical marker which correlates specifically with drug effects. The presence or quantity of a pharmacodynamic

20   marker is not related to the disease state or disorder for which the drug is being administered; therefore, the presence or quantity of the marker is indicative of the presence or activity of the drug in a subject. For example, a pharmacodynamic marker may be indicative of the concentration of the drug in a biological tissue, in that the marker is either expressed or transcribed or not expressed or transcribed in that tissue in relationship to the level of the

25   drug. In this fashion, the distribution or uptake of the drug may be monitored by the pharmacodynamic marker. Similarly, the presence or quantity of the pharmacodynamic marker may be related to the presence or quantity of the metabolic product of a drug, such that the presence or quantity of the marker is indicative of the relative breakdown rate of the drug *in vivo*. Pharmacodynamic markers are of particular use in increasing the sensitivity of

30   detection of drug effects, particularly when the drug is administered in low doses. Since even a small amount of a drug may be sufficient to activate multiple rounds of marker (e.g.,

an ARCAP marker) transcription or expression, the amplified marker may be in a quantity which is more readily detectable than the drug itself. Also, the marker may be more easily detected due to the nature of the marker itself; for example, using the methods described herein, anti-ARCAP antibodies may be employed in an immune-based detection system for

5   an ARCAP protein marker, or ARCAP-specific radiolabeled probes may be used to detect an ARCAP mRNA marker. Furthermore, the use of a pharmacodynamic marker may offer mechanism-based prediction of risk due to drug treatment beyond the range of possible direct observations. Examples of the use of pharmacodynamic markers in the art are described in Matsuda *et al.* US 6,033,862; Hattis *et al.* (1991) Env. Health Perspect. 90: 229-238;

10   Schentag (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S21-S24; and Nicolau (1999) *Am, J. Health-Syst. Pharm.* 56 Suppl. 3: S16-S20.

The ARCAP molecules of the invention are also useful as pharmacogenomic markers. As used herein, a "pharmacogenomic marker" is an objective biochemical marker which correlates with a specific clinical drug response or susceptibility in a subject (see, e.g.,

15   McLeod *et al.* (1999) *Eur. J. Cancer* 35:1650-1652). The presence or quantity of the pharmacogenomic marker is related to the predicted response of the subject to a specific drug or class of drugs prior to administration of the drug. By assessing the presence or quantity of one or more pharmacogenomic markers in a subject, a drug therapy which is most appropriate for the subject, or which is predicted to have a greater degree of success, may be

20   selected. For example, based on the presence or quantity of RNA, or protein (e.g., ARCAP protein or RNA) for specific tumor markers in a subject, a drug or course of treatment may be selected which is optimized for the treatment of the specific tumor likely to be present in the subject. Similarly, the presence or absence of a specific sequence mutation in ARCAP DNA may correlate with ARCAP drug response. The use of pharmacogenomic markers

25   therefore permits the application of the most appropriate treatment for each subject without having to administer the therapy.

Pharmacogenomics

The ARCAP molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on ARCAP activity (e.g., ARCAP gene

30   expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) ARCAP associated disorders (e.g.,

liver cancer) associated with aberrant or unwanted ARCAP activity. In conjunction with such treatment, pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by

5   altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer an ARCAP molecule or ARCAP modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with an ARCAP molecule or ARCAP modulator.

10   Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, for example, Eichelbaum, M. *et al.* (1996) *Clin. Exp. Pharmacol. Physiol.* 23:983-985 and Linder, M.W. *et al.* (1997) *Clin. Chem.* 43:254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single

15   factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical

20   complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

       One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association," relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (e.g., a "bi-allelic" gene

25   marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants). Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be

30   generated from a combination of some ten-million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in

a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP may be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular

5   pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

Alternatively, a method termed the "candidate gene approach", can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a

10   drug's target is known (e.g., an ARCAP protein of the present invention), all common variants of that gene can be fairly easily identified in the population, and it can be determined if having one version of the gene versus another is associated with a particular drug response.

Alternatively, a method termed the "gene expression profiling", can be utilized to identify genes that predict drug response. For example, the gene expression of an animal

15   dosed with a drug (e.g., an ARCAP molecule or ARCAP modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment of an individual. This knowledge, when applied to

20   dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with an ARCAP molecule or ARCAP modulator, such as a modulator identified by one of the exemplary screening assays described herein.

The present invention further provides methods for identifying new agents, or

25   combinations, that are based on identifying agents that modulate the activity of one or more of the gene products encoded by one or more of the ARCAP genes of the present invention, where these products may be associated with resistance of the cells to a therapeutic agent. Specifically, the activity of the proteins encoded by the ARCAP genes of the present invention can be used as a basis for identifying agents for overcoming agent resistance. By

30   blocking the activity of one or more of the resistance proteins, target cells, e.g., human cells,

20

will become sensitive to treatment with an agent that the unmodified target cells were resistant to.

Monitoring the influence of agents (e.g., drugs) on the expression or activity of an ARCAP protein can be applied in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase ARCAP gene expression, protein levels, an ARCAP activity can be monitored in clinical trials of subjects exhibiting decreased ARCAP gene expression, protein levels, or ARCAP activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease ARCAP gene expression, protein levels, or an ARCAP activity can be monitored in clinical trials of subjects exhibiting increased ARCAP gene expression, protein levels, or ARCAP activity. In such clinical trials, the expression or activity of an ARCAP gene, and preferably other genes that have been implicated in, for example, an ARCAP-associated disorder can be used as a "read out" or marker of the phenotype of a particular cell.

Without further elaboration, it is believed that one skilled in the art can, based on the above disclosure and the example below, utilize the present invention to its fullest extent. The following examples are to be construed as merely illustrative of how one skilled in the art can isolate and use the polypeptides and nucleic acids of the invention, and are not limitative of the remainder of the disclosure in any way. Any publications cited in this disclosure are hereby incorporated by reference.

<u>Example</u>

Materials and Methods

Patient samples. Hepatocellular carcinoma patients from the Department of Surgery, Veterans General Hospital, Taipei, Taiwan, were recruited into this study. Diagnosis of hepatoma was made by sonography, angiography, computer tomography, and/or magnetic resonance imaging. Clinical information for each patient was recorded, including the patients' age, sex, serum alpha-fetoprotein level, liver function, tumor size, tumor location, pathological staging, disease-free interval, time of recurrence, and location of tumor recurrence. Informed consent was obtained from each patient. For each liver cancer patient, tissue was harvested from the tumor, as well as from the normal liver tissue that was adjacent to the tumor.

RNA extraction and reverse transcription for complement DNA. Tissue specimens were frozen immediately after surgical resection and stored at -80°C. RNA was extracted using acid guanidinium thiocyanate and phenol/chloroform extraction as described in Chomczynski et al. (1987) *Anal. Biochem.* 162:156-159. Approximately 0.5 g of frozen

5   tissue was homogenized with 4 ml RNAzol B (Biotecx Laboratories, Houston, Texas) using a polytron. DNA was sheared using a Douncer with a type-B pestle. After adding 0.4 ml of chloroform, vigorous vortexing, and standing on ice for 5 minutes, the mixture was separated by centrifugation at 12,000 g at 4°C for 15 minutes. The upper aqueous phase containing total RNA was precipitated with an equal volume of isopropanol.

10   Complement DNA was synthesized from the 1 µg of total RNA. Reverse transcription was performed in a volume of 30 µl, containing RNA and 1 x first strand buffer (10 mM DTT, 500 µM dNTPs, 50 ng/ml oligo-dT, and 100 units MMLV reverse transcriptase) at 37°C for 1 hour (Life Technologies). The samples were then denatured at 95°C for 5 minutes.

15   *PCR amplification.* cDNA (1 µl) was PCR amplified in a volume of 25 µl containing 0.8 µM of primers, 50 µM of each dNTP (Takara), 1x PCR buffer, and 1.25 units Taq Polymerase (Pharmacia). As a control for cDNA quality, a test PCR reaction was carried out using the human transferrin gene primers Tref8 (GGAACATTTTGGCAAAGACA [SEQ ID NO:4]; derived from nucleotides 971-990 of the transferring cDNA sequence) and Tref9

20   (ATTCATGATCTT(C/T)GCGATGC [SEQ ID NO:5]; derived from nucleotides 1307-1288 of the transferrin cDNA sequence). These primer sequences were chosen form the eighth and ninth exons of the human transferrin gene, respectively. PCR was performed using an initial denaturation at 94°C for 10 minutes; followed by 30 cycles of 94°C for 1 minutes, 55°C for 1 minute, and 72°C for 1 minute; then a final extension at 72°C for 10 minutes. A

25   successful PCR yielded a 336 bp product, indicating that the cDNA template is at least 1.4 kb from the poly(A) end (nt 2362).

The steroid receptor superfamily clones were generated by amino acid-based homology PCR using degenerate primers encoding highly conserved sequence motifs in the zinc finger of DNA binding domain of steroid receptors. The primers encoded the amino

30   acid sequences DYSTGYHY (SEQ ID NO:6), CKXFFKR (SEQ ID NO:7), and CPACRFXKC (SEQ ID NO:8), all of which are described in Maksymowych et al (1992)

Receptor 2:225-240. The forward and reverse primer sequences were
RCAYTTIIIIARICKRCAIKMNKGRCA (SEQ ID NO:11), and
GAYRARKCIWCIGGIWRICAYT (SEQ ID NO:12). PCR was performed under low
stringency annealing/extension conditions (42°C/65°C) for 5 cycles, before amplifying the

5   templates at high-stringency conditions (55°C/72°C). The PCR products were subcloned
into the TA vector (Invitrogen), and the resulting plasmid used to transform DH5α cells.
Clones were randomly picked and checked for an insert size of about 170 bp, a length
corresponding to a zinc finger. A high frequency of clones contained appropriately sized
inserts (85-90%).

10   *DNA Sequence Analysis of novel ARCAP.* Positive clones were picked and sequenced
using Applied Biosystems model 377 DNA sequencers. The cloned sequences were
analyzed using the BLASTN program (Zehetner et al. (1994) *Nature* 367:489-491). Partial
cDNA clones bearing sequence similarity to members of the steroid receptor superfamily
were selected for further study. Of these clones, one clone named ARCAP, was

15   characterized and fully cloned as described herein.

        Isolation of full length clone. To obtain the complete open reading frame of ARCAP,
a commercial cDNA library of G2 hepatoma cell line (Clontech) was probed with the
ARCAP partial clone. In addition, hepatoma cDNA libraries were constructed from RNA
isolated from a hepatoma tumor and normal human liver tissue, both donated by a male

20   patient that died from trauma. Approximately 5 μg of mRNA was used for reverse
transcription. The cDNA library was prepared using the lambda ZAP II system (Stratagene).
The library was amplified from 1.1×10^6 PFU of primary recombinant clones. The average
insert size was 1.2 kb. More than 95 % of the clones are recombinants.

        To isolate novel clones from the liver cDNA libraries, radio-labeled probes were

25   prepared using a PCR reaction supplemented with a ^{32}P-labeled dCTP. Specifically, a
labeling reaction of 100 μl volume contained 1 x buffer (1.5 μM $MgCl_2$, 0.5 μl Taq
polymerase (25 units/ml), 200 μM each of dGTP, dTTP, and dATP, and 25-50 μM dCTP)
and 5 μl of ^{32}P-α-dCTP. The PCR conditions for producing labeled probes was essentially
the same as described above for amplifying steroid receptor clones based on conserved

30   domains.

The cDNA library was screened at a moderate density (16,000 PFU per 150 mm plate). Twenty plates were initially screened. Pre-hybridization was carried out in 5 x SSC, 2 x Denhardt's, 100 mg/ml single-stranded salmon sperm DNA, and 0.1% SDS at 55°C. Hybridization was performed in the same solution but supplemented with 1 x 10^7 cpm/ml

5   solution of denatured PCR-labeled probes. After incubation at 55°C for 20 hours, low stringency washing was performed for 60 minutes at room temperature using 2 x SSC, 0.1% SDS. The blots were visualized by autoradiography (24 hours exposure) using Kodak X-AR film and one intensifying screen (Lighting Plus).

Cloning by 5'-RACE and 3'-RACE. RACE is a procedure for amplification of a

10   cDNA template between a known internal site and unknown sequences at either 5' or the 3' ends (Fronhman et al., (1988) *Proc. Natl. Acad. Sci. USA* 85:8998-9001). The basic protocol performed in this study was described in the literature accompanying Clontech's RACE kits.

GSP1 (TCTGGTGGTTGCACTGAGCT; SEQ ID NO:13) and GSP2 (ACAATGTCAGCTCAGGCTC; SEQ ID NO:14) primers were designed in-house and

15   based on the sequence of ARCAP. Human hepatoma cell line G2 mRNA was used as template for synthesis of first strand cDNA. This synthesis was performed using primers 3'-CDS (AAGCAGTGGTAACAACGCAGAGTACT$_{30}$ NN; SEQ ID NO:15) for 3' RACE or Smart-oligo (T$_{25}$NN; SEQ ID NO:16) plus 5'-CDS (AAGCAGTGGTAACAACGCAGAG TACGCGGG; SEQ ID NO:17) for 5' RACE.

20   After synthesis of first strand cDNA , 5' RACE was performed by PCR using Smart and GSP1 primers. For 3' RACE, PCR was performed using GSP2 and 3'CDS primers. Each fragment after PCR were cloning into pGEM-easy vector (Promega). After screening, the correct clones were picked, and plasmid DNAs were purified. The full length ARCAP cDNA sequence was thus obtained.

25   *Production of ARCAP antibody.* The ARCAP clone was digested with EcoR1 and inserted into pGEX2T to produce a ARCAP-GST expression plasmid (Pharmacia). This plasmid was used to transform BL21 bacteria, which were then induced using IPTG to express the fusion protein. The fusion protein was purified from bacterial lysate using Glutathione Sepharose 4B affinity chromatography. The ARCAP protein was detected by

30   Western blotting using an anti-GST antibody. Twenty micrograms of fusion protein was injected into Balb/c mice to raise ARCAP antiserum.

ARCAP expression analysis. To determine the mRNA expression pattern for ARCAP, RNA samples prepared from tissues and cell lines were analyzed by Northern blot. Twenty micrograms each of total RNA from liver, fetal brain, four cell lines derived from liver (HepG2, Hep3B, VGH/22T, VGH/59T), blood cell lines (K562, U932, Ramos, Jurkat),

5   and HeLa cells were separated on a formaldehyde gel and transferred to a filter membrane. Filter hybridization was carried out with the labeled probes at 42°C in 5 x SSC, 5 x Denhardt's, 5 mg/ml denatured salmon sperm DNA, 50% formamide, and 0.1% SDS. Washing was first performed at room temperature in 2 x SSC/0.1% SDS for 30 minutes, followed by washing at 42°C in 2 x SSC/0.1% SDS for 30 minutes, and then a final washing

10   at 55°C in 0.2 x SSC/0.1% SDS for 30 minutes. Autoradiography film was exposed using one intensifying screen for 72 hours.

For in situ hybridization, liver cancer tissue sections were obtained from the patient samples described above. A riboprobe of ARCAP was prepared using the partial ARCAP clone and a biotin labeling kit (NEN). Hybridization and washes were carried out according

15   to established protocols. For ARCAP protein expression, an antibody was raised using a GST-ARCAP fusion expression vector as described herein. This antibody was then used to identify the presence of ARCAP protein in the various tissue sections.

ARCAP transactivation of androgen-responsive genes. COS-1 cells were cultured for at least 48 hours before transfection in Dulbecco's modified Eagle's medium (Life

20   Technologies, Inc.) supplemented with 10% fetal calf serum that had been stripped of steroids by treatment with dextran-coated charcoal. Cells were grown to 60-80% confluence, washed, removed, and seeded at a density of 5×10^4 cells/well in fresh medium in 35 mm wells (Corning) of a microtitre tissue culture plate. After 20 hours, cells were washed once with serum-free medium and transfected using Fugene 6 (Roche) according to the

25   manufacturer's directions. Ten hours thereafter, cells were washed twice with the appropriate medium and incubated in 2 ml of medium containing steroid or vehicle. After 48 hours, cells were recovered and assayed for luciferase activity according to the manufacturer's instructions (Promega).

Luciferase activity was corrected for the corresponding β-galactosidase activity to

30   give relative activity. β-Galactosidase assays were performed in a 96-well plate (Corning) as follows: 10 μl of sample extract were incubated with 80 μl of buffer Z and 10 μl of

o-nitrophenyl-β-D-galactopyranoside (4 mg/ml) at 30°C for 2 hours. The reaction was terminated by the addition of 50 μl of 1 M Na_2CO_3. A_{420} values were obtained using a MR5000 plate reader (Dynatech), and activity was calculated as described herein. Transfections were performed in triplicate and repeated at least three times.

5   Hepatoma A2 cells were subcultured in DMEM medium (Life Technologies, Inc.) supplemented with 10% fetal calf serum. Cells were seeded at 1×10^5 cells/well in 35 mm wells at least 24 hours before transfection. For studies examining the interaction between AR and ARCAP-1, cells were cultured in DMEM supplemented with 10% dextran-coated charcoal-fetal calf serum. Cells were transfected using Fugene 6 for a period of 8 hours,

10   washed, and cultured in medium containing vehicle or steroid. Cells were harvested at 48 hours after transfection and analyzed as described above.

A 450 base fragment containing an androgen-responsive element (TGGGTACATTTTGTTC; SEQ ID NO:9) from a portion of the human alpha-fetoprotein promoter 4385 bp from the ATG start coding site (Genbank ID: G178242) was found to have

15   transactivation response ability when cloned into a heterologous gene. This fragment was amplified by PCR from human genomic DNA (Clontech). A element androgen responsive element containing the sequence TGGGTA<u>GG</u>TTTTG<u>C</u>TC (SEQ ID NO:10) was made by site direct mutagenesis. The mutated nucleotides are indicated by underlining. This product was cloned into the pGEM-easy vector (Promega), and the sequence of the element was

20   confirmed by sequencing. The inserted DNA were subcloned into the appropriate sites in the luciferase reporter plasmid pGL3basic (Promega).

AR and ARCAP were fused in frame to the amino terminus of enhanced GFP (EGFP) in the vector pEGFP-N1 (Clontech) using an EcoRI site. This plasmid was used to transform into *E. coli* (DH5). Positive clones were identified using mini-plasmid preparation and

25   restriction enzyme analysis. Selected constructs were confirmed by sequencing. COS7 and A2 cells were cultured in DMEM supplemented with 10% fetal calf serum, penicillin/streptomycin (100 μg/ml), and L-glutamine (2 mM). Subconfluent monolayers (10^6 cells in 100 mm dishes) were transfected with 10 μg of AR and ARCAP cDNA by calcium phosphate. Thirty hours later, fresh media were replaced with medium

30   supplemented with DHT (10 nM). The cells were observed using a fluorescence microscope (Nikon).

Immunoprecipitations. Expression plasmids encoding androgen receptor, ARA70, and ARCAP-HA or ARCAP-Xpress epitope fusion protein were produced in pCDNA III (Invitrogen) using EcoRI. After transfection into COS7 cells, protein was labeled using the in vitro-coupled transcription and translation kit, T7-TNT (Promega), for incorporating

5   ^{35}S-methione. At 48 hours post-transfection, the cultures were supplemented to 10 nM DHT. Cells were lysed, and the lysate was incubated with 1 ml of immunoprecipitation buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 0.2 mM Na_3VO_4, 0.5% Nonidet P-40, 1 mM phenylmethylsulfonyl fluoride, 1 mM dithiothreitol, 25 μg/ml leupeptin, 25 μg/ml aprotinin, and 25 μg/ml pepstatin). The lysate was then mixed and incubated with AR, HA, or Xpress

10   antibody (Invitrogen) on ice for 60 minutes, at which time 10 μl of protein A-sepharose beads (Pharmacia) were added. The immunoprecipitations were incubated for 16 hours at 4°C with continuous mixing. Immunoprecipitated complexes were collected by centrifugation at 2000 rpm at 4°C for 10 minutes. The pelleted beads were washed three times and mixed with SDS sample buffer. Samples were resolved on 8% polyacrylamide

15   gels at 200 V for 45 minutes. The gel was fixed for 30 minutes in 10% propanol and 10% acetic acid, soaked in Amplify (Amersham Pharmacia Biotech) for 30 minutes, dried under vacuum, and exposed to X-ray film for 4 to 24 hours at 70°C.

Yeast two hybrid analysis. For independent cloning of ARCAP, androgen receptor was used as a bait to screen the matchmaker yeast two hybrid library (Clontech). Briefly, the

20   full length insert of androgen receptor sequence was cloned into pS2-1 (Clontech), and the resulting construct used to screen a yeast-two hybrid testis library constructed in pACT2 (Clontech), using the yeast host Y187, according to the manufacturer's protocols. Two hundred thousand transformed cells were plated onto media lacking tryptophan, leucine, and adenine, but supplement with 1 μM dihydrotestosterone (DHT). Adenine-positive, LacZ-

25   positive, histidine-positive colonies were isolated, and plasmid DNA obtained therefrom. The plasmids were transformed into *E. coli*.

pACT2 plasmids containing cDNA were identified by colony PCR using primers specific for the LEU2 gene present in pACT2. Specificity of interaction for with the human androgen receptor (hAR) was determined by examining the liquid LacZ activity of cDNA

30   clones in the presence of the GAL4DBD:hAR fusion protein versus the activity observed in

the presence of the GAL4DBD alone. After sequencing and examination of the GenBank database, it was determined that one interacting clone was encoded by ARCAP.

The yeast two-hybrid system was also actively used to confirm the interaction of ARCAP with hAR. pAS2-constructs were co-transformed with pACT2-ARCAP or pACT2

5   alone into yeast host Y187 according to the manufacturer's protocol (Clontech). Trp-positive, Leu-positive colonies were inoculated in triplicate onto selective media and grown at 30°C overnight in the presence or absence of 1 μM DHT. Samples were diluted to an A_{600} of 0.2 and re-grown to an A_{600} of 0.6-0.8. Samples were divided into three aliquots of 1 ml each. Cells were recovered by centrifugation at 14,000 rpm for 5 minutes, washed once with

10   buffer Z (0.1 M sodium phosphate, pH 7.0, 10 mM KCl, and 10 mM $MgSO_4$), and resuspended in 800 μl of buffer Z containing 21 μl of 2-mercaptoethanol. Then 10 μl of 0.1% SDS were added, followed by the addition of 50 μl of chloroform. Samples were vortexed for 1 minute and incubated at 30°C, and then 200 μl of o-nitrophenyl-β-D-galactopyranoside (4 mg/ml in buffer Z) were added. Reactions were timed and terminated

15   upon observing an obvious yellow color or after 1 hour by the addition of 500 μl of 1 M Na_2CO_3. A_{420} of the samples was determined, and activity was calculated as follows: (A_{420} x 1000)/(A_{600} x time). All assays were performed in triplicate and repeated at least three times.

Results

The full length ARCAP cDNA and the protein it encodes is described above.

20   ARCAP mRNA levels were assessed in normal human tissues and in various cell lines, including hepatoma cell lines, using Northern blotting. ARCAP was detected only in normal human heart and skeletal muscle tissues. High levels of ARCAP expression was detected in hepatoma cells including 3B, 22T, Huh 7, G2, and A2.

Paired liver tumor and normal tissue adjacent to the tumor were isolated from 40 liver

25   cancer patients. Using Northern blotting, it was discovered that ARCAP mRNA was generally highly expressed in the tumor but expressed very little, if at all, in the adjacent normal liver tissue. Using ARCAP-specific antibodies, the presence of ARCAP protein in tumor tissue and the general absence of ARCAP protein in the adjacent normal tissue was confirmed. In situ hybridization of ARCAP mRNA in human hepatoma tissues also

30   indicated that ARCAP mRNA was abundant in tumor cells but rare in normal cells. These

data concerning the expression profile of ARCAP mRNA and protein indicated that ARCAP expression is a marker for liver cancer.

To determine cellular localization of ARCAP protein, the ARCAP coding region was fused to GFP in an expression vector, and the vector transfected into human hepatoma A2

5 cells. The fusion protein was localized in the nucleus, indicating that ARCAP is a nuclear protein.

A yeast two hybrid system was used to determine whether ARCAP binds to androgen receptor. A full length androgen receptor was cloned adjacent to a GAL4 DNA binding domain and used as a bait to screen for His3-positives clones from the human matchmaker

10 liver library (Clontech). The interaction of AR and ARCAP in vivo was confirmed by this study. To further confirm that ARCAP binds to the androgen receptor, ARCAP and the androgen receptor were co-expressed as fusion proteins. Since immunoprecipitation of androgen receptor led to isolation of ARCAP and immunoprecipitation of ARCAP led to isolation of androgen receptor, the physical interaction of ARCAP and androgen receptor

15 was confirmed.

To determine whether the physical association between the androgen receptor and ARCAP protein was biochemically significant, the α-fetoprotein gene promoter was used in a luciferase reporter construct. The α-fetoprotein (AFP) gene is a model system for the studying developmental control of gene expression. AFP is expressed at a high level in fetal

20 liver, but its transcription declines rapidly after birth and is hardly detectable in adult life. However, the AFP gene is often reactivated to a high level when hepatomas or teratomas develop (Shulman et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:8288-8292). An enhancer region in the AFP promoter contains an androgen responsive element (ARE). A control isogenic luciferase reporter containing a mutated ARE was also used in the study to

25 determine whether the ARE was responsible for mediating any biochemical effects.

G2 cells were transfected with the wild type or control reporter and (1) mock DNA, (2) DNA encoding androgen receptor, (3) DNA encoding ARCAP, or (4) DNA encoding androgen receptor and DNA encoding ARCAP. Little change in luciferase activity was observed when the control reporter was used. However, using the wild type reporter,

30 luciferase activity was increased at least 2 x (relative to the control reporter) when androgen receptor was expressed. Surprisingly, the simultaneous expression of androgen receptor and